

# Appendix

## Appendix A1.1 Study Characteristics: Barnett, Jung, Yarosz, Thomas, Hornbeck, Stechuk, & Burns, 2008 (randomized controlled trial)

Characteristic	Description
<b>Study citation</b>	Barnett, W., Jung, K., Yarosz, D., Thomas, J., Hornbeck, A., Stechuk, R., & Burns, S. (2008). Educational effects of the <i>Tools of the Mind</i> curriculum: a randomized trial. <i>Early Childhood Research Quarterly, 23</i> (3), 299–313.
<b>Participants</b>	In one school selected for the study, 7 classrooms on one floor were available for <i>Tools of the Mind</i> implementation and 11 classrooms on another floor were available for the control condition. Teachers and assistants were randomly assigned to classrooms using a stratified assignment procedure, and then three- and four-year-old children were randomly assigned to either <i>Tools of the Mind</i> curriculum classrooms or district curriculum classrooms. Poverty level, achievement, and minority status were similar across intervention and comparison groups. Among the children sampled, 93 percent are Hispanic, and about 70 percent consider Spanish their primary home language. Although the overall student attrition rate was more than 25 percent, and student consent after random assignment led to differential attrition, the post-attrition intervention and comparison samples were equivalent on achievement pretests. After one year, 85 <i>Tools of the Mind</i> students and 117 comparison students remained in the sample.
<b>Setting</b>	This study was conducted in 18 classrooms in a low-income urban school with state-financed Abbott full-day preschool education.
<b>Intervention</b>	<i>Tools of the Mind</i> aims to aid learning and development while emphasizing emergent literacy and self-regulation. The two main goals of the curriculum are to develop underlying cognitive skills (such as self-regulation, deliberate memory, and focused attention) and to develop specific academic skills (such as symbolic thought, literacy, and an understanding of math). Play is the leading activity for developing such skills and the curriculum emphasizes the teacher's role in supporting the development of mature intentional dramatic play. The study was conducted during the first year of program implementation of <i>Tools of the Mind</i> .
<b>Comparison</b>	Control classrooms implemented the standard district-created curriculum, which was described as a full-day PreK balanced literacy curriculum with themes. In structured observations of the control group, frequently observed activities were art projects that correlated with the “letter of the week,” free play, large group movement and/or music, and such large group activities as story time. According to the study authors, although the control curriculum covered much of the same academic content and topics as <i>Tools of the Mind</i> , there was greater emphasis on teacher-imposed control and less on children's self-regulation.
<b>Primary outcomes and measurement</b>	For both pre- and post-tests, the authors administered Peabody Picture Vocabulary Test-III, Expressive One-Word Picture Vocabulary Test-Revised, Animal Pegs Subtest of the Wechsler Preschool Primary Scale of Intelligence, and two subtests of the Woodcock-Johnson-Revised test (Applied Problems and Letter-Word Identification). Get Ready to Read! screening tool was used only at post-test assessment. IDEA Oral Language Proficiency Test was administered for the subsample of Spanish-speaking children. Problem Behaviors Scale of the Social Skills Rating System was also used in the study, but not included in this report because it was outside the scope of the Early Childhood Education review. For a more detailed description of these outcome measures, see Appendix A2.1–2.4.
<b>Teacher and staff training</b>	Teachers assigned to the <i>Tools of the Mind</i> group received four full days of curriculum training before the start of the school year. During the school year, they received 30-minute classroom visits approximately once a week from a <i>Tools of the Mind</i> trainer to address any difficulties they were having with the curriculum. In addition, <i>Tools of the Mind</i> teachers received 1 half-day workshop and 5 one-hour lunchtime meetings to discuss aspects of the curriculum. Control group teachers received similar amounts of training. They attended workshops on the already established district curriculum given by the district for the same amount of time.

## Appendix A2.1 Outcome measures for the oral language domain

Outcome measure	Description
<b>Peabody Picture Vocabulary Test-III</b>	A standardized measure of children's receptive vocabulary that requires children to identify pictures that correspond to spoken words (as cited in Barnett et al., 2008).
<b>Expressive One-Word Picture Vocabulary Test-Revised</b>	A standardized measure of children's expressive vocabulary that requires them to name pictures of common objects, actions, and concepts (as cited in Barnett et al., 2008).
<b>IDEA Oral Language Proficiency Test</b>	This test assesses the receptive and expressive Spanish language skills of Spanish-speaking children (as cited in Barnett et al., 2008).

## Appendix A2.2 Outcome measures for the print knowledge domain

Outcome measure	Description
<b>Woodcock-Johnson-Revised: Letter-Word Identification subtest</b>	A standardized measure of children's ability to name printed letters and words (as cited in Barnett et al., 2008).
<b>Get Ready to Read!</b>	A nonstandardized measure of readiness for reading instruction focusing on three core domains (print knowledge, emergent writing skills, and linguistic awareness) across 20 items to which children indicate their response by pointing (as cited in Barnett et al., 2008).

## Appendix A2.3 Outcome measures for the cognition domain

Outcome measure	Description
<b>Wechsler Preschool Primary Scale of Intelligence: Animal Pegs subtest</b>	A subset from a standardized measure that assesses a child's nonverbal problem-solving and visual-motor proficiency as they place pegs of correct colors in a series of holes under pictures of animals (as cited in Barnett et al., 2008).

## Appendix A2.4 Outcome measures for the math domain

Outcome measure	Description
<b>Woodcock-Johnson-Revised: Applied Problems subtest</b>	A subtest from a standardized measure that assesses children's math skills by asking children to count small sets and to solve simple addition and subtraction questions using pictures (as cited in Barnett et al., 2008).

## Appendix A3.1 Summary of study findings included in the rating for the oral language domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>3</sup> ( <i>Tools of the Mind</i> – comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
			<i>Tools of the Mind</i> group	Comparison group				
<b>Barnett et al. (2008) (randomized controlled trial)<sup>7</sup></b>								
Peabody Picture Vocabulary Test-III (PPVT-III)	Three- to four-year-olds	18/198	nr (19.19)	nr (15.90)	3.59	0.21	ns	+8
Expressive One-Word Picture Vocabulary Test-Revised (EOWPVT-R)	Three- to four-year-olds	18/193	nr (14.06)	nr (12.22)	1.19	0.09	ns	+4
<b>Average for oral language<sup>8</sup></b>						<b>0.15</b>	<b>ns</b>	<b>+6</b>

ns = not statistically significant

nr = not reported

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the oral language domain. Subgroup findings for children who consider Spanish their primary language are not included in these ratings but are reported in Appendix A4.1.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. The standard deviations were provided by the author at the WWC request.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean difference is the hierarchical linear model (HLM) coefficient for the intervention's effect provided by the author at the WWC request.
4. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). For the formulas the WWC used to calculate statistical significance, see [Technical Details of WWC-Conducted Computations](#). For Barnett et al. (2008), no corrections for clustering or multiple comparisons were needed because the study reported findings were based on hierarchical linear model (HLM) analyses and were not statistically significant.
8. This row provides the study average, which is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

## Appendix A3.2 Summary of study findings included in the rating for the print knowledge domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study			WWC calculations		
			Mean outcome (standard deviation <sup>2</sup> )		Mean difference <sup>3</sup> ( <i>Tools of the Mind</i> – comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
			<i>Tools of the Mind</i> group	Comparison group				
<b>Barnett et al. (2008) (randomized controlled trial)<sup>7</sup></b>								
Woodcock-Johnson-Revised Letter-Word Identification subtest	Three- to four-year-olds	18/202	nr (12.87)	nr (11.92)	–0.45	–0.04	ns	–1
Get Ready to Read!	Three- to four-year-olds	18/220	nr (3.90)	nr (3.91)	0.13	0.03	ns	+1
<b>Average for print knowledge<sup>8</sup></b>						<b>0.00</b>	<b>ns</b>	<b>0</b>

ns = not statistically significant

nr = not reported

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the print knowledge domain.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. The standard deviations were provided by the author at the WWC request.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean difference is the hierarchical linear model (HLM) coefficient for the intervention's effect provided by the author at the WWC request.
4. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). For the formulas the WWC used to calculate statistical significance, see [Technical Details of WWC-Conducted Computations](#). For Barnett et al. (2008), no corrections for clustering or multiple comparisons were needed because the study reported findings were based on hierarchical linear model (HLM) analyses and were not statistically significant.
8. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

### Appendix A3.3 Summary of study findings included in the rating for the cognition domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study			WWC calculations		
			Mean outcome (standard deviation <sup>2</sup> )		Mean difference <sup>3</sup> ( <i>Tools of the Mind</i> – comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
			<i>Tools of the Mind</i> group	Comparison group				
<b>Barnett et al. (2008) (randomized controlled trial)<sup>7</sup></b>								
Wechsler Preschool Primary Scale of Intelligence (WPPSI) Animal Pegs Subtest	Three- to four-year-olds	18/200	nr (15.22)	nr (16.35)	0.84	0.05	ns	+2
<b>Average for cognition<sup>8</sup></b>						<b>0.05</b>	<b>ns</b>	<b>+2</b>

ns = not statistically significant

nr = not reported

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the cognition domain.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. The standard deviations were provided by the author at the WWC request.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean difference is the hierarchical linear model (HLM) coefficient for the intervention's effect provided by the author at the WWC request.
4. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). For the formulas the WWC used to calculate statistical significance, see [Technical Details of WWC-Conducted Computations](#). For Barnett et al. (2008), no corrections for clustering or multiple comparisons were needed because the study reported findings were based on hierarchical linear model (HLM) analyses and were not statistically significant.
8. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

## Appendix A3.4 Summary of study findings included in the rating for the math domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study			WWC calculations		
			Mean outcome (standard deviation <sup>2</sup> )		Mean difference <sup>3</sup> ( <i>Tools of the Mind</i> – comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
			<i>Tools of the Mind</i> group	Comparison group				
<b>Barnett et al. (2008) (randomized controlled trial)<sup>7</sup></b>								
Woodcock-Johnson-Revised Applied Problems	Three- to four-year-olds	18/202	nr (16.19)	nr (18.86)	3.07	0.17	ns	+7
<b>Average for math<sup>8</sup></b>						<b>0.17</b>	<b>ns</b>	<b>+7</b>

ns = not statistically significant

nr = not reported

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the math domain.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. The standard deviations were provided by the author at the WWC request.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean difference is the hierarchical linear model (HLM) coefficient for the intervention's effect provided by the author at the WWC request.
4. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). For the formulas the WWC used to calculate statistical significance, see [Technical Details of WWC-Conducted Computations](#). For Barnett et al. (2008), no corrections for clustering or multiple comparisons were needed because the study reported findings were based on hierarchical linear model (HLM) analyses and were not statistically significant.
8. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

## Appendix A4.1 Summary of subgroup findings by age for the oral language domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation <sup>2</sup> )		Mean difference <sup>3</sup> ( <i>Tools of the Mind</i> – comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
			<i>Tools of the Mind</i> group	Comparison group				
<b>Barnett et al. (2008) (randomized controlled trial)<sup>7</sup></b>								
IDEA Oral Language Proficiency Test in Spanish	Three- to four-year-olds	18/132	nr (8.49)	nr (6.82)	2.36	0.31	ns	+12

ns = not statistically significant

nr = not reported

1. This appendix presents subgroup findings for measures that fall in the oral language domain. The Oral Language Proficiency Test in Spanish was administered to children who considered Spanish their primary language (approximately 70 percent of the sample) to assess their Spanish language development. Total group scores were used for rating purposes and are presented in Appendix A3.1.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. The standard deviations were provided by the authors at the WWC request.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean difference is the hierarchical linear model (HLM) coefficient for the intervention's effect provided by the author at the WWC request.
4. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). For the formulas the WWC used to calculate statistical significance, see [Technical Details of WWC-Conducted Computations](#). For Barnett et al. (2008), no corrections for clustering were needed because the study reported findings were based on hierarchical linear model (HLM) analyses.

## Appendix A5.1 *Tools of the Mind* rating for the oral language domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.<sup>1</sup>

For the outcome domain of oral language, the WWC rated *Tools of the Mind* as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects because no studies showed statistically significant or substantively important effects, either positive or negative.

### Rating received

**No discernible effects:** No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative*.

**Met.** One study examined effects on oral language and did not show statistically significant or substantively important effects, either positive or negative.

### Other ratings considered

**Positive effects:** Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a *strong* design.

**Not met.** No study showed statistically significant positive effects.

#### AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

**Met.** No study showed statistically significant or substantively important negative effects.

**Potentially positive effects:** Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

**Not met.** No study showed a statistically significant or substantively important positive effect.

#### AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

**Not met.** No study showed a statistically significant or substantively important negative effect, but one study showed indeterminate effects.

**Mixed effects:** Evidence of inconsistent effects as demonstrated through EITHER of the following.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

**Not met.** No study showed a statistically significant or substantively important effect, either positive or negative.

#### OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an indeterminate effect than showing a statistically significant or substantively important effect.

**Not met.** No study showed a statistically significant or substantively important effect, while one study showed indeterminate effects.

(continued)

## Appendix A5.1 *Tools of the Mind* rating for the oral language domain (continued)

**Potentially negative effects:** Evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *negative* effect.

**Not met.** No study showed a statistically significant or substantively important negative effect.

**AND**

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, OR more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

**Met.** No study showed a statistically significant or substantively important positive effect.

**Negative effects:** Strong evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *negative* effects, at least one of which met WWC evidence standards for a *strong* design.

**Not met.** No study showed a statistically significant or substantively important negative effect.

**AND**

- Criterion 2: No studies showing statistically significant or substantively important *positive* effects.

**Met.** No study showed statistically significant or substantively important positive effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the [WWC Intervention Rating Scheme](#).

## Appendix A5.2 Tools of the Mind rating for the print knowledge domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.<sup>1</sup>

For the outcome domain of print knowledge, the WWC rated *Tools of the Mind* as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects because no studies showed statistically significant or substantively important effects, either positive or negative.

### Rating received

**No discernible effects:** No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative*.

**Met.** One study examined effects on print knowledge and did not show statistically significant or substantively important effects, either positive or negative.

### Other ratings considered

**Positive effects:** Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a *strong* design.

**Not met.** No study showed statistically significant positive effects.

#### AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

**Met.** No study showed statistically significant or substantively important negative effects.

**Potentially positive effects:** Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

**Not met.** No study showed a statistically significant or substantively important positive effect.

#### AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

**Not met.** No study showed a statistically significant or substantively important negative effect, but one study showed indeterminate effects.

**Mixed effects:** Evidence of inconsistent effects as demonstrated through EITHER of the following.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

**Not met.** No study showed a statistically significant or substantively important effect, either positive or negative.

#### OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an indeterminate effect than showing a statistically significant or substantively important effect.

**Not met.** No study showed a statistically significant or substantively important effect, while one study showed indeterminate effects.

(continued)

## Appendix A5.2 *Tools of the Mind* rating for the print knowledge domain (continued)

**Potentially negative effects:** Evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *negative* effect.

**Not met.** No study showed a statistically significant or substantively important negative effect.

**AND**

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, OR more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

**Met.** No study showed a statistically significant or substantively important positive effect.

**Negative effects:** Strong evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *negative* effects, at least one of which met WWC evidence standards for a *strong* design.

**Not met.** No study showed a statistically significant or substantively important negative effect.

**AND**

- Criterion 2: No studies showing statistically significant or substantively important *positive* effects.

**Met.** No study showed statistically significant or substantively important positive effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the [WWC Intervention Rating Scheme](#).

## Appendix A5.3 *Tools of the Mind* rating for the cognition domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.<sup>1</sup>

For the outcome domain of cognition, the WWC rated *Tools of the Mind* as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects because no studies showed statistically significant or substantively important effects, either positive or negative.

### Rating received

**No discernible effects:** No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative*.

**Met.** One study examined effects on cognition and did not show statistically significant or substantively important effects, either positive or negative.

### Other ratings considered

**Positive effects:** Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a *strong* design.

**Not met.** No study showed statistically significant positive effects.

#### AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

**Met.** No study showed statistically significant or substantively important negative effects.

**Potentially positive effects:** Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

**Not met.** No study showed a statistically significant or substantively important positive effect.

#### AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

**Not met.** No study showed a statistically significant or substantively important negative effect, but one study showed indeterminate effects.

**Mixed effects:** Evidence of inconsistent effects as demonstrated through EITHER of the following.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

**Not met.** No study showed a statistically significant or substantively important effect, either positive or negative.

#### OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an indeterminate effect than showing a statistically significant or substantively important effect.

**Not met.** No study showed a statistically significant or substantively important effect, while one study showed indeterminate effects.

(continued)

### Appendix A5.3 *Tools of the Mind* rating for the cognition domain (continued)

**Potentially negative effects:** Evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *negative* effect.

**Not met.** No study showed a statistically significant or substantively important negative effect.

**AND**

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, OR more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

**Met.** No study showed a statistically significant or substantively important positive effect.

**Negative effects:** Strong evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *negative* effects, at least one of which met WWC evidence standards for a *strong* design.

**Not met.** No study showed a statistically significant or substantively important negative effect.

**AND**

- Criterion 2: No studies showing statistically significant or substantively important *positive* effects.

**Met.** No study showed statistically significant or substantively important positive effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the [WWC Intervention Rating Scheme](#).

## Appendix A5.4 *Tools of the Mind* rating for the math domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.<sup>1</sup>

For the outcome domain of math, the WWC rated *Tools of the Mind* as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects because no studies showed statistically significant or substantively important effects, either positive or negative.

### Rating received

**No discernible effects:** No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative*.

**Met.** One study examined effects on math and did not show statistically significant or substantively important effects, either positive or negative.

### Other ratings considered

**Positive effects:** Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a *strong* design.

**Not met.** No study showed statistically significant positive effects.

#### AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

**Met.** No study showed statistically significant or substantively important negative effects.

**Potentially positive effects:** Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

**Not met.** No study showed a statistically significant or substantively important positive effect.

#### AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

**Not met.** No study showed a statistically significant or substantively important negative effect, but one study showed indeterminate effects.

**Mixed effects:** Evidence of inconsistent effects as demonstrated through EITHER of the following.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

**Not met.** No study showed a statistically significant or substantively important effect, either positive or negative.

#### OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an indeterminate effect than showing a statistically significant or substantively important effect.

**Not met.** No study showed a statistically significant or substantively important effect, while one study showed indeterminate effects.

(continued)

## Appendix A5.4 *Tools of the Mind* rating for the math domain (continued)

**Potentially negative effects:** Evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *negative* effect.

**Not met.** No study showed a statistically significant or substantively important negative effect.

**AND**

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, OR more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

**Met.** No study showed a statistically significant or substantively important positive effect.

**Negative effects:** Strong evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *negative* effects, at least one of which met WWC evidence standards for a *strong* design.

**Not met.** No study showed a statistically significant or substantively important negative effect.

**AND**

- Criterion 2: No studies showing statistically significant or substantively important *positive* effects.

**Met.** No study showed statistically significant or substantively important positive effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the [WWC Intervention Rating Scheme](#).

## Appendix A6    Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence <sup>1</sup>
		Schools	Students	
Oral language	1	1	198	Small
Print knowledge	1	1	220	Small
Cognition	1	1	200	Small
Math skills	1	1	202	Small
Phonological processing	0	na	na	na
Early reading/writing	0	na	na	na

na = not applicable/not studied

1. A rating of “medium to large” requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is “small.”